I know that I know nothing

Voelkel, J.G.; Brandt, M.J.; Colombo, M.

Published in:
Comprehensive Results in Social Psychology

DOI:
10.1080/23743603.2018.1464881

Publication date:
2018

Document Version
Peer reviewed version

Running head: I KNOW THAT I KNOW NOTHING

I know that I know nothing: Can puncturing the illusion of explanatory depth overcome the relationship between attitudinal dissimilarity and prejudice?

Jan G. Voelkel[a]

Mark J. Brandt[a1]

Matteo Colombo[b]

[a]: Department of Social Psychology, Tilburg University, Tilburg, the Netherlands

[b]: Tilburg Center for Logic, Ethics and Philosophy of Science, Tilburg University, Tilburg, the Netherlands

Status: Accepted (before proofing) at *Comprehensive Results in Social Psychology*

Email addresses: jangvoelkel.research@gmail.com, m.j.brandt@tilburguniversity.edu, m.colombo@tilburguniversity.edu

[1] Corresponding author. Email: m.j.brandt@tilburguniversity.edu

Abstract

People are prejudiced towards groups they perceive as having a worldview dissimilar from their own. This link between perceived attitudinal dissimilarity and prejudice is so stable that it has been described as a psychological law (Byrne, 1969). The current research tests whether reducing people's (over-)confidence in their own understanding of policies by puncturing their illusion of explanatory depth in the political domain will reduce the link between perceived attitudinal dissimilarity and prejudice. In an initial pre-registered experiment (N = 296), we did not find support for our hypothesis, but exploratory analyses indicated that the hypothesized effect occurred for political moderates (but not for people who identified as strong liberals/conservatives). However, despite successfully manipulating people's understanding of policies, in the main study (N = 492) we did not replicate the result of the initial experiment. We suggest potential explanations for our results and discuss future directions for research on breaking the link between attitudinal dissimilarity and prejudice.

Keywords: dissimilarity; prejudice; illusion of explanatory depth; political psychology; attitudes

I know that I know nothing - Can puncturing the illusion of explanatory depth overcome the relationship between attitudinal dissimilarity and prejudice?

"We only have one remaining bigotry. We don't want to be around anybody who disagrees with us" (Bill Clinton, 2014)

As former US president Bill Clinton recognized, people are prejudiced towards groups characterized by a worldview they see as different from their own (Brandt, 2017; Brandt, Reyna, Chambers, Crawford, & Wetherell, 2014; Chambers, Schlenker, & Collison, 2013; Crawford & Pilanski, 2014). Indeed, this link between perceiving a group as having dissimilar social and political attitudes, and expressions of prejudice (i.e. negative evaluations based on group membership[2], Crandall, Eshleman, & O'Brien, 2002) is so robust that is has been referred to as a psychological "law" (Byrne, 1969; Byrne & Nelson, 1965). Even people who are open to new experiences are prejudiced towards social groups they perceive as holding dissimilar social and political beliefs (Brandt, Chambers, Crawford, Wetherell, & Reyna, 2015). These prejudices are consequential, as they prevent resolutions of socially and culturally divisive issues, and can lead to discriminatory and aggressive behavior (Cox & Devine, 2014; cf. Chambers & Melnyk, 2006; Wynn, 2016).

Despite its social importance, developing interventions that break or reduce the link between attitudinal dissimilarity and prejudice is challenging. While social psychologists have identified a number of ways to reduce prejudice – for example, work inspired by the contact hypothesis (Pettigrew & Tropp, 2006) – these methods have often focused on improving specific

[2] This definition is a well-accepted definition of prejudice (Brandt & Proulx, 2016; Brown, 2010; Crandall, Ferguson, & Bahns, 2013; Stangor, 2009). While it is focused on the central psychological feature of prejudice (i.e. negative affect), this definition does not say anything about the appropriateness or justifiability of any given prejudice.

intergroup relationships by (1) highlighting the similarities between groups or (2) emphasizing the distinctiveness of the group while creating commonalities between groups (Dovidio & Gaertner, 1999). First, contact between outgroup members can lead to less prejudice by transforming the cognitive representations of outgroup members as being closer to the in-group or closer to the self, and, thus, more similar to the self (Cameron, Rutland, Brown, & Douch, 2006; Gaertner, Mann, Dovidio, & Murrell, 1990; cf. Stathi, Cameron, Hartley, & Bradford, 2014). Second, research has indicated that distinctiveness between subgroups can actually foster more positive intergroup attitudes under certain conditions (Hornsey & Hogg, 2000). For example, multiculturalism explicitly marks group differences as positive, but emphasizes simultaneously how the different groups can contribute to a common larger group (e.g. society; Park & Judd, 2005).

What is not clear is how to break the link between attitudinal dissimilarity and prejudice without directly altering the perceived attitudinal dissimilarity and without creating a common group-identity. This kind of intervention is important, because creating attitudinal similarity in some domains is not possible, or at the very least difficult (e.g., political differences are characterized by worldview dissimilarity and it is difficult to change political beliefs). Similarly, creating a common goal (e.g. contributing to society) is not straightforward in the political context, because liberals and conservatives often have fundamentally different ideas about how and what groups should contribute to society (e.g. Graham, Haidt, & Nosek, 2009). The goal of the current study was to test whether the link between perceived attitudinal dissimilarity and prejudice can be weakened or even overcome when people are reminded that they do not fully understand policies related to their political attitudes.

Attitudinal Dissimilarity, Prejudices, and the Illusion of Explanatory Depth

It is well known that people express prejudice towards others with different attitudes and opinions compared to their own. Existing literature suggests that this effect occurs for two basic reasons. The first is as part of humans' coalition detection system (Amodio, 2014; Boyer, Firat, & van Leeuwen, 2015; Kurzban, Tooby, & Cosmides, 2001). According to this research, people have an evolved capacity for identifying friends and foes. This capacity relies on social cues. These can include ethnicity, but also political beliefs. People use these cues automatically to identify who belongs to which group (Pietraszewski, Curry, Petersen, Cosmides, & Tooby, 2015). People who are perceived as holding dissimilar political beliefs are then flagged as foes and, thus, met with prejudice. The second is a reaction to incongruity (Brandt et al, 2014, 2015; Byrne, 1969; Crawford, 2014; Schaller, Boyd, Yohannes, & O'Brien, 1995). People desire to understand the world and encountering people whose attitudes and opinions violate that understanding frustrates that goal. In response, people derogate those with different attitudes and values in an attempt to bolster the validity of their own way of viewing the world.

A third possibility, that we consider here, is that people are overconfident about the validity/correctness of their attitudes, opinions, and worldviews. People who are overconfident in their own opinions are then more likely to believe that others with different attitudes are wrong. Thus, people who are overconfident think that they have reasons to be prejudiced towards others with dissimilar views. In general, people who are confident about their attitudes are often less likely to change their attitudes in the face of disconfirming evidence (Bassili, 1996; Tormala & Petty, 2002). However, when it comes to politics, it is not just that people are confident in their attitudes, but they are also overly confident with regard to their understanding of policies and how the policies will work (Fernbach, Rogers, Fox, & Sloman, 2013). This is an instance of the illusion of explanatory depth, where "people feel they understand the world with far greater detail, coherence, and depth than they really do" (Rozenblit & Keil, 2002, p. 522). When this

illusion of explanatory depth is punctured, however, by asking people to explain in detail how a particular public policy works, people's levels of political extremity drop (Fernbach et al., 2013). In short, explaining the mechanism underlying public policies is very difficult and often makes people realize that they do not understand the policy as well as they thought. This reduces their confidence in their understanding of these policies and the extremity of their attitudes.

Bringing together ideas and methods from existing literature on the illusion of explanatory depth, and on the link between attitudinal dissimilarity and prejudice, we aimed to test the following hypothesis: Reducing people's (over-)confidence in their own understanding of policies by puncturing their illusion of explanatory depth in the political domain should reduce the link between attitudinal dissimilarity and prejudice. This intervention is unique in that it aims to reduce prejudice towards dissimilar groups without changing the view that the other group has a different worldview and/or different goals for society, but by making people more willing to accept this worldview dissimilarity.

Initial Study

We conducted an initial pre-registered experiment to test this idea (see supplementary materials for details).[3] In short, we combined the methods used to examine the link between attitudinal dissimilarity and prejudice (Brandt et al., 2015) with the methods used to understand explanatory depth and extreme political attitudes (Fernbach et al., 2013; see also Rozenblit & Keil 2002). Participants rated how dissimilar 20 groups were in terms of holding social and political attitudes different from their own, described either their reasons for their position on three political policies (control condition) or the causal mechanisms behind the three policies

[3] The pre-registration (including hypotheses and analysis plan), the materials, the data, and the analysis script for the initial study can be found here: https://osf.io/nbrww/ (pre-registration) and https://osf.io/wksmy/ (data and analysis script). Minor errors in the pre-registered analysis script were identified and corrected during performing the actual analysis (the differences are described here: https://osf.io/78knx/).

(experimental condition), and then completed prejudice measures for the same 20 groups. We hypothesized that participants in the experimental condition would have a weaker association between perceived attitudinal dissimilarity and prejudice than people in the control condition. Our hypothesis was not supported. Exploratory analyses, however, identified a more nuanced picture. We found that the experimental condition – where we punctured the illusion of explanatory depth – reduced the attitudinal dissimilarity-prejudice association for political moderates, but not for participants identifying as conservative or liberal. Here we aimed to dig into this potential moderator.[4]

Potential Moderator: Ideological Extremity

Ideological extremity vs. moderation may moderate the effect of our manipulation for at least two reasons. First, political moderates are less likely to see their positions as superior (Brandt, Evans, & Crawford, 2015; Toner, Leary, Asher, & Jongman-Sereno, 2013) and to moralize political issues (Ryan, 2014). Moral convictions are a strong predictor of hostility towards others who disagree with them (Skitka, Bauman, & Sargis, 2005) and they are also resistant to change and the influence of others (Aramovich, Lytle, & Skitka, 2012; Skitka, Bauman, & Lytle, 2009; Wisneski, Lytle, & Skitka, 2009; see also Colombo, Bucher, & Inbar, 2016). Therefore, it may be easier to change moderates' minds because the moral barrier that any manipulation of the illusion of explanatory depth needs to overcome is likely to be less strong for people with a moderate political ideology than for people with a more extreme ideology.

Second, moderates are more tolerant of political ambiguity than political extremists are. That is, they have less problems to accept compromises and are more willing to remain

[4] Our initial study also indicated that the experimental condition might even increase the strength of the dissimilarity-prejudice relationship for strong conservatives (but not for moderate conservatives or liberals). However, we think that this finding is less likely to be replicated as (a) we do not have good theoretical reasons to believe that this effect should occur and (b) the number of strong conservatives in both conditions was very low (9 strong conservatives in each writing task condition). Nonetheless, the current design allows us to test whether this effect replicates.

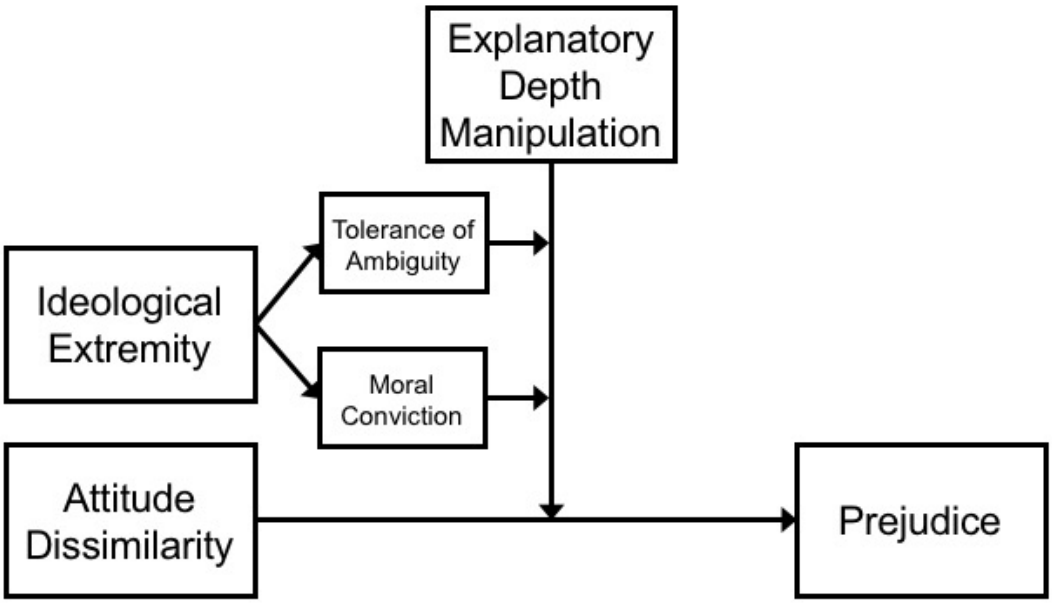
undecided or to accept multiple perspectives to the same issue (McClosky & Chong, 1985). Moderates also see the political world in less stark, black and white terms (Lammers, Koch, Conway, & Brandt, 2016; Tetlock, 1984) than people with a more extreme political ideology do. Being confronted with one's lack of understanding of policies and the related thought that then one's prejudices towards dissimilar groups may be less justified is likely to create ambiguity; it weakens the positions that one has held so far and creates more possibilities for alternative solutions. If moderates are better able to accept this ambiguity, they may be willing to update their previous beliefs and, thus, might be more likely to change their prejudices compared to political extremists.

The Current Research

Figure 1 summarizes the model. The main goal of the current research was to test if puncturing the illusion of explanatory depth reduces the association between perceived attitudinal dissimilarity and prejudice for political moderates. In order to test this prediction, participants were randomly assigned to one of two writing task conditions (explaining the mechanisms underlying different policies versus generating reasons for one's position on the policies; using the same instruction as in the initial study and as in Fernbach et al., 2013). We predicted that the positive association between perceived attitudinal dissimilarity and prejudice towards social groups would be smaller in the mechanism explanation condition than in the reason generation condition because the mechanism explanation task would puncture the illusion of explanatory depth. However, we expected that this effect would only occur for political moderates, but not for people with a more extreme political ideology.

An additional goal of this research was to examine potential mechanisms underlying this effect. We predicted that the moderation effect of ideological extremity might be driven by two features that distinguish moderates from the political extreme: (a) their weaker inclination to

Figure 1: *Three-way Interaction Effect of Perceived Attitudinal Dissimilarity, Manipulation of Explanatory Depth, and Extremity of Political Ideology on Prejudice as Mediated by Strength of Moral Convictions and Tolerance of Political Ambiguity*



moralize political issues and (b) their higher acceptance of political ambiguity. Therefore, our design included measures for both strength of moral convictions and tolerance of political ambiguity to examine these mediation predictions. The pre-registration (including hypotheses and analysis plan), the materials, the data, and the analysis script for this study can be found here: https://osf.io/etwq9/ (pre-registration) and https://osf.io/29qmy/ (data and analysis script). Minor errors in the pre-registered analysis script were identified and corrected during performing the actual analysis (the differences are described here: https://osf.io/tgxec/).

Method

Participants

Participants were recruited from across the Unites States using Amazon Mechanical Turk and received a payment for their participation ($1/participant). We conducted a power analysis[5]

[5] We thank Jake Westfall for his advice on the best way to conduct the power analysis.

with PANGEA (Westfall, 2016). This analysis suggested that we should collect about 500 participants who rate dissimilarity and prejudice for 30 social groups. This sample size gives us approximately 83% power to detect a small effect size (d = 0.20) and approaching 100 % power to detect the mean effect size in social psychology (d = 0.45; Westfall, 2016, based on Richard, Bond, & Stokes-Zoota, 2003). A more detailed description for our power analysis rationale can be found in the supplementary materials.

Our initial sample size was n = 808. Participants with the same IP address (n = 69) were removed from the data set (we kept only the first case), as were people who did not complete the study (n = 247).[6] Of these 247 participants, 1 person exited the study before they were assigned to a condition. When participants were assigned to a condition, significantly more participants did not complete the survey in the mechanism explanation condition than in the reason generation condition (41 % vs. 25 %), $\chi^2(1) = 20.75$, $p < .001$. This finding indicates that the dropout rate in the mechanism explanation condition was significantly higher than in the reason generation condition (further discussed in the Discussion section). Our final sample consisted of 492 participants (275 in the reason generation condition and 217 in the mechanism explanation condition), of which 274 people identified as male, 215 as female, and 3 as other. The average age amounted to 36.03 (SD = 11.71). While we prepared the manuscript for re-submission and double-checked the fit with the preregistration, we realized the discrepancy between our final sample size of n = 492 and our statement in the preregistration that we would collect participants until we reached a sample size of n = 500. However, given a time difference of roughly 2.5 months between data collection and the detection of this difference, we decided in consultation with the Associate Editor to proceed with the given sample size of n = 492.

[6] Not completing the study means here that a participant does not have a single valid response for at least one of our variables of interest.

Procedure

We used a between-subjects design with two different conditions (writing task: explanation of mechanisms vs. generation of reasons).[7] For all participants, the experiment consisted of five parts. In the first part, participants rated the perceived attitudinal dissimilarity of 30 groups. In the second part, participants completed measures of the two potential mediators: moral convictions and tolerance of ambiguity. The third part consisted of a writing task, in which participants were randomly assigned to either describe the mechanisms behind three policies in as much detail as possible (mechanism explanation condition) or to generate reasons for their position on three policies (reason generation condition). In the fourth part, participants' completed measures of prejudice towards the 30 groups. In the final part, participants were asked demographic questions, including a measure of their political ideology, the proposed moderator in our analysis. Afterwards, participants were debriefed and thanked for their participation.

Materials

We aimed to test the relationship between attitudinal dissimilarity and prejudice for a representative sample of groups that was balanced with regard to the perceived political ideology of the groups (a list of these groups is available in the Appendix). In order to do so, we used 30 frequently named social groups in the US (cf. Koch, Imhoff, Dotsch, Unkelbach, & Alves, 2016). We started by including the 30 most frequently named social groups (see Koch et al., 2016, Table 1). However, this collection of groups was slightly liberally skewed (M = 55.11 using the ratings provided by Koch et al. on a conservative-liberal-scale from 0 to 100). Therefore, we replaced the

[7] In the initial study, we also had a factor that manipulated the order in which the writing task and the dissimilarity ratings took place (see supplementary materials). In one condition, the writing task followed the dissimilarity ratings. In another condition, the writing task preceded the dissimilarity ratings. We included this factor initially in order to examine whether the explanatory depth manipulation significantly influences participants' dissimilarity ratings. We found it did not. This finding also has implications for the possibility that a common rater bias may influence our findings. As the writing task can be taxing in terms of time and thinking, the lack of a significant difference between the two order conditions suggests that having time and thought between the two types of ratings does not significantly affect the results. Therefore, we dropped this order manipulation from the current study.

least frequently named from these 30 social groups, which are liberal (≥ 60 rating on Koch et al.'s 0-100 scale), by the most frequently named non-chosen social groups, which are conservative (≤ 40 rating on Koch et al.'s 0-100 scale), until the mean perceived ideology of the groups was balanced (M = 49.87). Specifically, we replaced goths with preps, children with elderly, teachers with business people, and gays with white collar. Thus, the groups we used were both diverse and balanced with regard to their perceived political ideology.

Independent Variable: Perceived Attitudinal Dissimilarity. Perceived attitudinal dissimilarity was measured for each target group with the following item: "Please indicate the extent to which you see each of the following groups as holding political or social beliefs different from your own". Participants answered on a slider scale from 1 (not at all different from me) to 7 (very different from me) with 4 as default (Brandt et al., 2015). The order of the groups was randomized for each participant.

Mediators: Strength of Moral Conviction and Tolerance of Political Ambiguity. We proposed two mediators for the moderation effect of political extremity on the association between perceived attitudinal dissimilarity and prejudice: the strength of moral convictions and tolerance of ambiguity. Moral convictions were measured for each of the three policies the participants selected with the following item: "How much are your feelings about the following policies connected to your core moral beliefs or convictions?" on a scale from 1 (not at all) to 7 (very much) (based on Skitka et al., 2005). The scores for the three items were averaged to form a strength of moral convictions composite such that higher values indicate stronger moral convictions. The mean on the recoded scale from 0 to 1 was 0.77 (SD = 0.18).

Intolerance of political ambiguity was measured with an adapted version of the Intolerance of Ambiguity scale (McClosky & Chong, 1985, OVS items in Table 6). We slightly changed the wording of the items to use rating scales rather than binary choices as the response

format. The adapted scale consisted of four items (e.g., "On important public issues, you should always keep in mind that there is more than one side to most issues", reverse-scored) answered on a scale from 1 (strongly disagree) to 7 (strongly agree). The scores for the four items were recoded and averaged to form an intolerance of political ambiguity composite such that higher values indicate more intolerance of political ambiguity. As the reliability coefficient was low for this scale (Cronbach's alpha = .39), the results involving intolerance of political ambiguity should be treated with caution. The mean on the recoded scale from 0 to 1 was 0.34 (*SD* = 0.16).

Moderator 1: Explanatory Depth Manipulation. With regard to the writing task factor, participants went through either a manipulation designed to decrease their self-rated understanding of policies, or through a control condition. Note that while our methodology is consistent with earlier work on the illusion of explanatory depth for artifacts (Rozenblit & Keil, 2002), we used the modified version designed by Fernbach and colleagues (2013), who focused on public policies. First, participants selected the three political issues that they find most important from a list of ten issues. The list of political issues included five of the policies used by Fernbach and colleagues and five additional policies that are relatively specific and of current relevance.[8] We included more policies than Fernbach and colleague and asked the participants to choose their three most important issues to ensure that participants realized that their lack of understanding of policies applies to topics they care about. Afterwards, participants rated their position on each of the three chosen issues on a scale from 1 (strongly against) to 7 (strongly in favor). In addition, participants rated their understanding of the policies on a scale from 1 (vague

[8] In the initial study, we used all six policies provided by Fernbach and colleagues (2013) and added four new policies. However, the results showed that for one of the six original policies and one of the newly included policies, there was no decrease in understanding at all in the mechanism explanation condition (cf. Table SM.1 in the supplementary materials). Therefore, these two policies were replaced by two new policies.

understanding) to 7 (thorough understanding). We used the original instructions from Fernbach et al. (2013) to explain the scale and the different levels of understanding.

Next, participants were randomly assigned to either a mechanism explanation writing task condition or a reason generation writing task condition. These conditions were nearly identical to the conditions used by Fernbach et al (2013), with only minor changes to accommodate the changes of our study (e.g., selection of three important issues). In the mechanism explanation condition, participants were asked to describe all the details they know about a policy, going from the first step to the last, and providing the causal connection between the steps. In the reason generation condition, participants were asked to write down all the reasons they have for their position on a policy, going from the most important to the least. In both conditions, participants went through the task for each of the three chosen policies and rated their understanding of the policy again directly afterwards using the same scale as above. The mechanism explanation writing task was expected to decrease participants' understanding ratings, while the reason generation writing task was not expected to affect the understanding ratings (Fernbach et al., 2013).

Dependent variable: Prejudice. Our dependent variable, prejudice, was measured with three items: disliking, preferred social distance to, and perceived immorality of the target groups.[9] The first two items reflect common measures of prejudices while addressing different components, namely affect and behavioral intentions (Brandt et al., 2015), while perceived immorality is an important dimension of the perception of others (Goodwin, Piazza, & Rozin, 2014; Skitka et al., 2005). We assessed disliking with a feeling thermometer slider rating on a scale from 0 (very cold, dislike quite a lot) to 100 (very warm, like quite a lot) with a default of

[9] In the initial study, we used a restriction of rights as third item (cf. Crawford, 2014; see supplementary materials). However, as this item was highly skewed and did not correlate strongly with the other two items, it was replaced in the current study.

50 (reverse scored). Social distance was measured with one item reading, "how willing would you be to occasionally spend social time with a person who is part of the following groups?", answered on a slider scale from 1 (not at all willing) to 7 (very willing) with a default of 4 (reverse scored) (Brandt et al., 2015). Perceived immorality was assessed with a similar slider rating as dislike, but with different anchors – ranging from 0 (ultimate evil) to 100 (ultimate good) with a default of 50 (reverse scored). The order of the three questions and the order of the groups for which the ratings were made were randomized. We specified *a priori* that, if the prejudice indicators were highly intercorrelated (Cronbach's alpha ≥ .7), the items would be combined into a prejudice composite. Otherwise, we would conduct separate analyses for the different items. As Cronbach's alpha was 0.83, the three items were averaged to form a prejudice composite.

Moderator 2: Extremity of Political Ideology. In the last part of the questionnaire, participants filled out a short demographic questionnaire including a measure of their political ideology. Political ideology was measured with a single item reading, "generally speaking, do you usually think of yourself as conservative, moderate, or liberal?", answered on a scale from -5 (very conservative) to 0 (moderate) to 5 (very liberal). Ideological extremity was then computed as the absolute value of this measure, so that it ranged from 0 (moderate) to 5 (extremist). In addition, participants' ideological type was computed as following: -1 for conservatives (< 0 on the ideology measure), 0 for moderates (= 0 on the ideology measure), 1 for liberals (> 0 on the ideology measure).[10] The sample was quite balanced with regard to ideological extremity (*M* =

[10] The trichotomy of the continuous ideology measure may initially seem like a strange choice. However, when we include the interaction between extremity and ideology type in our models, the interaction term recreates the continuous measure. This approach has the added benefit of separating out the effect of ideology type from ideology extremity and telling us if the effects of extremity have different effects for liberals and conservatives.

2.53, *SD* = 1.81), but included more liberals (*n* = 263) than moderates (*n* = 113) and conservatives (*n* = 116).

Covariates (for robustness check analysis). The demographic questionnaire also included measures of gender, age, and education, which were included as covariates in an additional robustness check analysis (see below). Gender was measured by asking participants to "please indicate your gender". Participants could identify as male, female, or other (and specify how they identified). Age was measured with a single item reading, "please indicate your age", on an open response scale. Education was measured with a single item reading, "what is the highest level of school that you have completed?" and five response categories: less than high school; high school diploma / GED; some college; Bachelor's degree; postgraduate (Master's degree, Ph.D., professional degree). The mean response amounted to 3.57 (*SD* = 0.88).

Results

We used multilevel analyses to test our hypotheses. For all multilevel analyses discussed in this paper, we used the lmer function of the lmerTest package in R (Kuznetsova, Brockhoff, & Christensen, 2016). Further, we used one-tailed tests to test all directional hypotheses that we specified a priori.

Preregistered Analyses

Manipulation Check: Change in Understanding Ratings (Preregistered). For our study, it is essential that the participants in the mechanism explanation condition (but not in the reason generation condition) realize that their understanding of the policies is worse than they expected. That is, we hypothesized to replicate Fernbach et al.'s finding that the mechanism explanation task leads to a reduced understanding of the policies compared to the reason generation task. To test this, we conducted a multilevel analysis with two different levels. Level 2 was constituted by the subjects and Level 1 was constituted by the issues. We conducted analyses

with three different dependent variables: a) the understanding rating at time 1 (higher ratings indicating better understanding), b) the difference score of the two understanding ratings after and before the manipulation (higher ratings indicating improved understanding), and c) the understanding rating at time 2 (higher ratings indicating better understanding). We included a fixed effect for writing task condition (reason generation task coded as -0.5, mechanism explanation task coded as 0.5) and a random intercept.

We did not expect any differences in the understanding ratings before the writing task manipulation. Indeed, the effect of writing task condition was not significant, $b = -0.06$, $SE = 0.10$, $t(490) = -0.58$, $p = .559$. This result indicates that there were no significant differences between the two conditions with regard to the understanding of the policies before the writing task.

We expected that understanding ratings would decrease in the mechanism explanation condition, but not in the reason generation condition. This prediction was supported by the data. The effect of writing task condition on change in understanding ratings was significant, $b = -0.32$, $SE = 0.07$, $t(490) = -4.53$, $p < .001$ (one-tailed). In the mechanism explanation condition, there was a significant decrease in understanding, $b = -0.24$, $SE = 0.05$, $t(490) = -4.65$, $p < .001$ (one-tailed), while in the reason generation condition, there was no significant change in understanding ratings, $b = 0.07$, $SE = 0.05$, $t(490) = 1.58$, $p = .114$. These results indicate that participants realized that their understanding was not as good as they thought in the mechanism explanation condition, but not in the reason generation condition.

The understanding ratings after the writing task manipulation were significantly lower in the mechanism explanation condition than in the reason generation condition, $b = -0.38$, $SE = 0.12$, $t(490) = -3.21$, $p = .001$. Thus, we can reasonably conclude that our manipulation of

reducing people's confidence in their understanding of political issues in the experimental condition was successful.

Main analysis: The effect of the explanatory depth manipulation on the attitudinal dissimilarity-prejudice association moderated by extremity of political ideology (Preregistered). The main hypothesis in our study is that puncturing the illusion of explanatory depth decreases the strength of the relationship between attitudinal dissimilarity and prejudice for moderates but not for people who identify strongly as liberal or conservative. That is, we predicted that, for political moderates, the positive association between attitudinal dissimilarity and prejudices will be smaller in the mechanism explanation writing task condition than in the reason generation writing task condition. Conversely, we did not expect a difference in the attitudinal dissimilarity-prejudice relationship between mechanism explanation writing task condition and reason generation writing task condition for people with an extreme political ideology.

In order to test our hypothesis, we conducted a multilevel analysis with two different levels. Level 2 was constituted by the subjects and Level 1 was constituted by the 30 social groups that we used for our measures of attitudinal dissimilarity and prejudices. We included the prejudice composite as dependent variable (rescaled to range from 0 to 1, higher ratings indicating more prejudice). In addition, we included fixed effects for (a) attitudinal dissimilarity (first rescaled to range from 0 to 1, then person mean-centered), (b) writing task condition (reason generation task coded as -0.5, mechanism explanation task coded as 0.5), (c) ideological extremity (the absolute value of the political ideology variable, ranging from 0 to 5, higher ratings indicating a more extreme political ideology), and (d) type of ideology (conservative coded as -1, moderate coded as 0, and liberal coded as 1). Further, we included fixed effects for

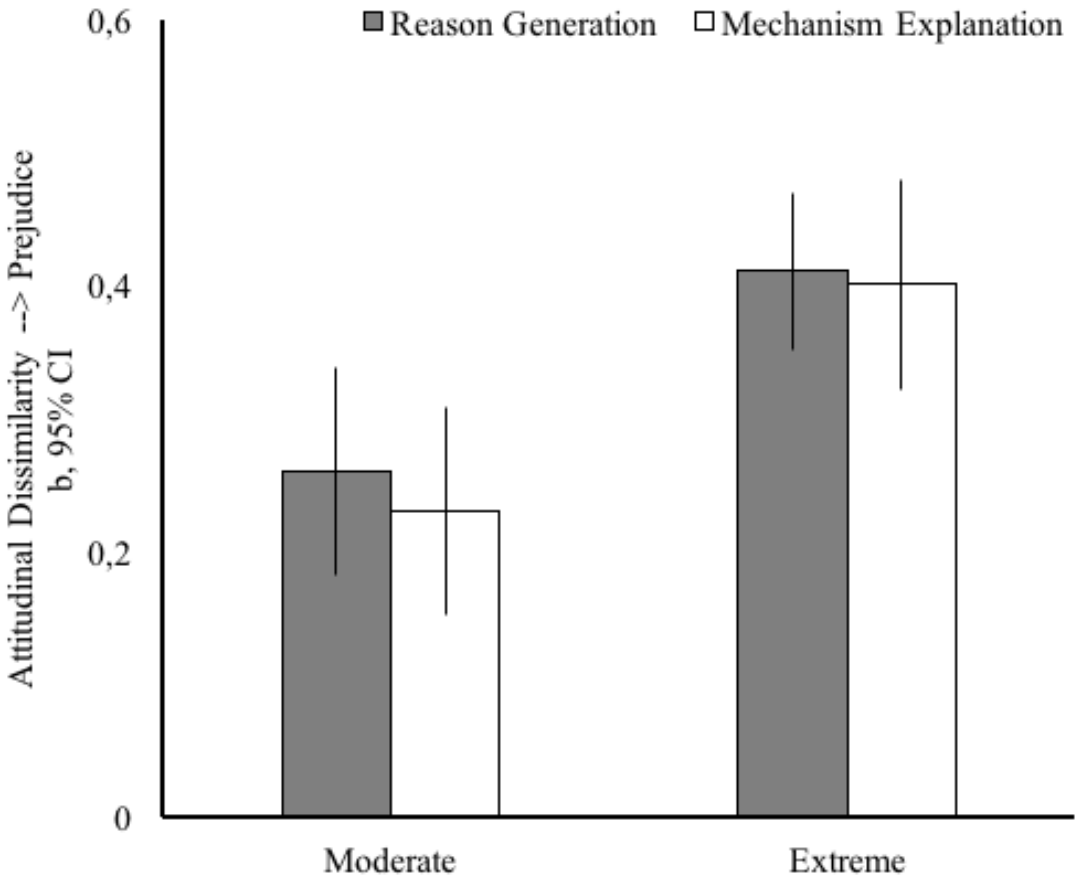
all two-, three- and four-way interactions between these variables and their interaction as well as a random intercept and a random slope for the effect of attitudinal dissimilarity.

The main effect of attitudinal dissimilarity was expected to be positive representing the robust finding of past research that individuals express more prejudices towards groups that are perceived as being more dissimilar. Our main hypothesis was then that an interaction effect of attitudinal dissimilarity, writing task condition, and ideological extremity emerges indicating that the effect of attitudinal dissimilarity is weaker in the mechanism explanation condition than in the reason generation condition for moderates but not for participants with an extreme political ideology. Notably, our analysis also allowed us to examine whether the results differ for liberal extremists and conservative extremists. This would be indicated by a significant four-way interaction effect. In the initial study, we found some evidence for such a pattern although we think it is less likely to be replicated (see footnote 3). Indeed, when we included a main effect of ideological type and its two-, three- and four-way interaction terms with the other three variables in our model, these effects were non-significant (all ps > .066). Thus, we removed these effects from our final model.

The analysis does not support our main hypothesis. The three-way interaction effect of attitudinal dissimilary, writing task condition, and ideological extremity was not significant, b = 0.01, SE = 0.02, $t(466) = 0.33$, $p = .371$ (one-tailed). As specified in our pre-registration, we nonetheless probed the effect of attitudinal dissimilarity for moderates and extremists in the two experimental conditions.

As can be seen in Figure 2, the analysis for moderates did not provide evidence for the effectiveness of the writing task condition in influencing the association between attitudinal

Figure 2: *The Unstandardized Beta (and 95% Confidence Intervals) of Perceived Attitudinal Dissimilarity on Prejudice Depending on Writing Task Condition and Ideological Extremity in the Main Study*



dissimilarity and prejudice. The two-way interaction effect of attitudinal dissimilarity and writing task condition was in the expected direction, but non-significant, b = -0.03, SE = 0.05, $t(480)$ = -0.57, p = .284 (one-tailed). For moderates in the reason generation condition, the effect of attitudinal dissimilarity was strong and significant, b = 0.26, SE = 0.04, $t(480)$ = 7.03, p < .001. For moderates in the mechanism explanation condition, the effect of attitudinal dissimilarity was also strong and significant, b = 0.23, SE = 0.04, $t(481)$ = 5.89, p < .001. It is clear from Figure 2 that these two simple effects are similar in size in both experimental conditions.

There was also no evidence for an effect of our writing task manipulation on the relationship between attitudinal dissimilarity and prejudice for political extremists. The two-way interaction effect of attitudinal dissimilarity and writing task condition was not significant, b = -0.00, SE = 0.05, $t(453)$ = -0.05, p = .957. For extremists in the reason generation condition, the effect of attitudinal dissimilarity was strong and significant, b = 0.41, SE = 0.03, $t(454)$ = 12.30, p < .001. For extremists in the mechanism explanation condition, the effect of attitudinal dissimilarity was also strong and significant, b = 0.40, SE = 0.04, $t(451)$ = 10.61, p < .001. It is clear from Figure 2 that these two simple effects are similar in size in both experimental conditions.

Mediation model (Preregistered). We hypothesized that the difference in the effectiveness of puncturing the illusion of explanatory depth between political moderates and extremists is driven by (a) the weaker moral convictions of political moderates and (b) the higher tolerance of political ambiguity of political moderates. Although our pre-registration stated that we would not conduct the mediation analyses without first finding the predicted three-way interaction, we decided to conduct the analyses to give a full presentation of the results.

In order to test these predictions, we carried out multilevel mediation analysis (Zhang, Zyphur, & Preacher, 2009). In the first step, we conducted two OLS regression analyses and (a) regressed strength of moral convictions (averaged across issues) on ideological extremity and (b) regressed intolerance of political ambiguity on ideological extremity. In the second step, we included the proposed mediators in the multilevel model that we used for our main analyses. For each mediator, we included its main effect as well as interaction effects with perceived attitudinal dissimilarity and the writing task condition.

With regard to the first step, our mediation hypotheses predicted that ideological extremity has a positive and significant effect on (a) strength of moral convictions and (b)

intolerance of political ambiguity. With regard to the second step, our mediation hypotheses predicted that the three-way interaction effect involving attitudinal dissimilarity, writing task condition, and extremity of ideology is reduced (and potentially become nonsignificant). In contrast, we expected that (a) the three-way interaction effect involving attitudinal dissimilarity, writing task condition, and strength of moral conviction and/or (b) the three-way interaction effect involving attitudinal dissimilarity, writing task condition, and intolerance of political ambiguity are significant. We expected that follow-up analysis show then that the effect of attitudinal dissimilarity is reduced in the mechanism explanation condition compared to the reason generation condition for people with weaker moral convictions and/or people who are less intolerant of political ambiguity. In contrast, we did not expect such an effect for people with stronger moral convictions and people who are more intolerant of political ambiguity. Significance tests for indirect effects were conducted with the Sobel test (Preacher & Leonardelli, 2001). Recent methodological research has suggested that indirect effect can be significant even if the total effect is not significant (Rucker, Preacher, Tormala, & Petty, 2011).

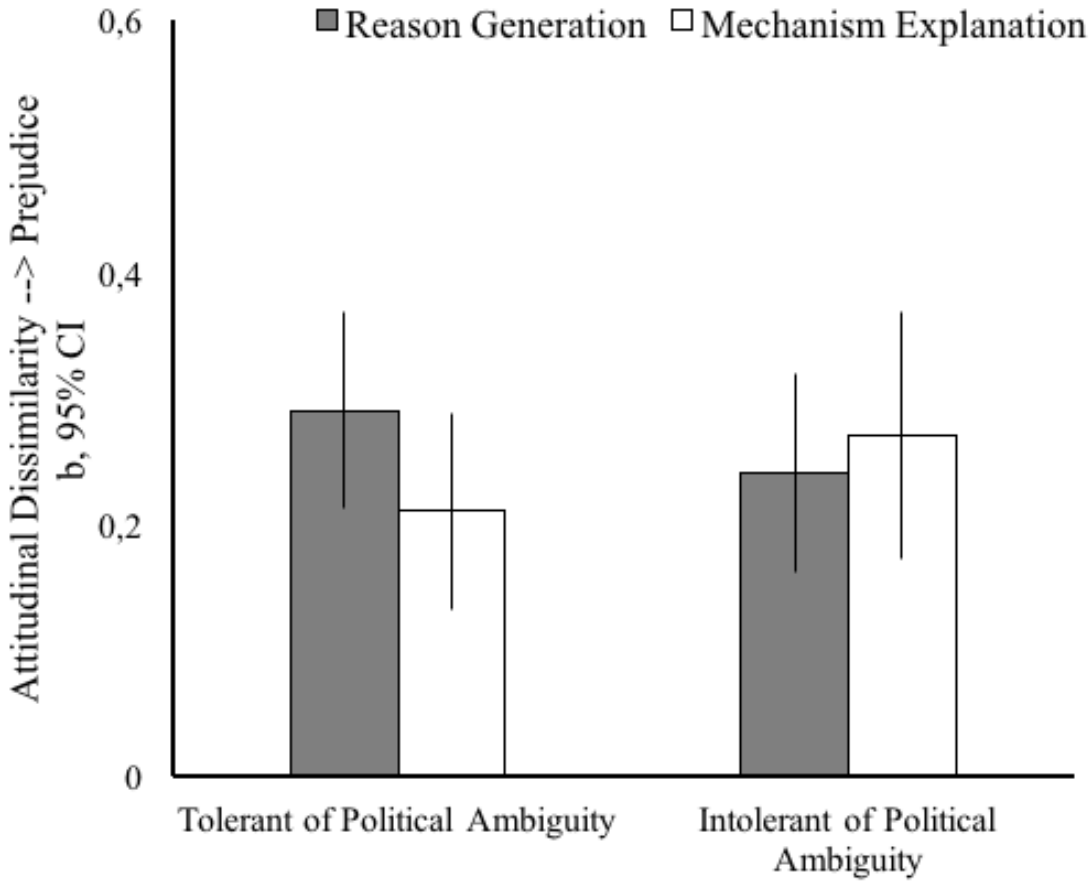
Step 1. As predicted, we found that ideological extremity was positively and significantly associated with participants' strength of moral convictions, $b = 0.02$, $SE = 0.00$, $t(490) = 4.70$, $p < .001$ (one-tailed). However, ideological extremity was not significantly related to intolerance of political ambiguity, $b = 0.01$, $SE = 0.00$, $t(490) = 1.51$, $p = .066$ (one-tailed).

Step 2. We did not find evidence that the attitudinal dissimilarity × writing task condition interaction effect depended on the strength of participants' moral convictions. The three-way interaction of attitudinal dissimilarity, writing task condition, and strength of moral convictions was not in the predicted direction and not significant, $b = -0.04$, $SE = 0.17$, $t(469) = -0.22$, $p = .587$ (one-tailed).

In contrast, we did find evidence that this interaction effect depended on participants' levels of intolerance of political ambiguity. The three-way interaction of attitudinal dissimilarity, writing task condition, and intolerance of political ambiguity was in the predicted direction and significant, $b = 0.38$, $SE = 0.19$, $t(458) = 2.02$, $p = .022$ (one-tailed). We probed the effect of attitudinal dissimilarity for participants who are relatively intolerant of political ambiguity (one standard deviation above the mean) and participants who are relatively tolerant of political ambiguity (one standard deviation below the mean) in the two experimental conditions.

The analysis for participants who are relatively intolerant of political ambiguity did not provide evidence for the effectiveness of the writing task condition in influencing the association between attitudinal dissimilarity and prejudice. The attitudinal dissimilarity × writing task condition interaction effect was non-significant, $b = 0.03$, $SE = 0.06$, $t(470) = 0.49$, $p = .623$. The size of the slopes are plotted in Figure 3. For participants who are relatively intolerant of political ambiguity in the reason generation condition, the effect of attitudinal dissimilarity was strong and significant, $b = 0.24$, $SE = 0.04$, $t(471) = 5.53$, $p < .001$. For participants who are relatively intolerant of political ambiguity in the mechanism explanation condition, the effect of attitudinal dissimilarity was also strong and significant, $b = 0.27$, $SE = 0.05$, $t(468) = 6.01$, $p < .001$. There was a non-significant trend for participants who are tolerant of political ambiguity that the relationship between attitudinal dissimilarity and prejudice was weaker in the mechanism explanation condition than in the reason generation condition. The attitudinal dissimilarity × writing task condition interaction effect was not significant, $b = -0.09$, $SE = 0.06$, $t(470) = -1.53$, $p = .063$ (one-tailed). For participants who are relatively tolerant of political ambiguity in the reason generation condition, the effect of attitudinal dissimilarity was strong and significant, $b = 0.29$, $SE = 0.04$, $t(468) = 7.35$, $p < .001$. For participants who are relatively tolerant of political ambiguity in the mechanism explanation condition, the effect of attitudinal dissimilarity was also

Figure 3: *The Unstandardized Beta (and 95% Confidence Intervals) of Perceived Attitudinal Dissimilarity on Prejudice Depending on Writing Task Condition and Intolerance of Political Ambiguity*



strong and significant, $b = 0.20$, $SE = 0.04$, $t(471) = 4.66$, $p < .001$. Figure 3 plots these two simple effects of dissimilarity. Although the three-way interaction was significant and in the predicted direction, the subsequent two-way interaction effect did not surpass our pre-registered alpha level of .05 (one-tailed). In total, the hypothesis is not supported.

Indirect effects. Both indirect effects were non-significant, Sobel test for mediation via strength of moral convictions: $z = -0.22$, $p = .587$ (one-tailed); Sobel test for mediation via intolerance of political ambiguity: $z = 1.21$, $p = .113$ (one-tailed). The direct effect, that is the three-way interaction of attitudinal dissimilarity, writing task condition, and ideological

extremity, remained not significant in the mediation model, $b = 0.01$, $SE = 0.02$, $t(462) = 0.36$, $p = .715$.

Additional analyses: Differences in dissimilarity ratings (Preregistered). We tested whether political ideology and/or ideological extremity affect how dissimilar participants view the 30 social groups. In the initial study, we did not find such an effect (see supplementary materials). In the proposed study, we also tested for such an effect. Therefore, we conducted a multilevel analysis with two different levels. Level 2 was constituted by the subjects and Level 1 was constituted by the social groups that we used for our measures of attitudinal dissimilarity and prejudices. We included the attitudinal dissimilarity variable as dependent variable (rescaled to range from 0 to 1, higher ratings indicating higher perceived attitudinal dissimilarity). In addition, we included fixed effects for (a) ideological extremity (the absolute value of the political ideology variable, ranging from 0 to 5, higher ratings indicating a more extreme political ideology), and (b) type of ideology (conservative coded as -1, moderate coded as 0, and liberal coded as 1). Further, we included a fixed effect for the two -way interaction between these variables as well as a random intercept. We expected to find no significant main or interaction effects.

We found a weak but significant effect of ideological extremity on perceived attitudinal dissimilarity, $b = -0.01$, $SE = 0.00$, $t(489) = -2.16$, $p = .031$. Surprisingly, participants with a more extreme ideology perceived the social groups as less dissimilar overall. The interaction effect and the main effect of ideological type were non-significant (both ps $> .266$).

Robustness check (Preregistered). We examined the robustness of the results of the main and mediation analyses by including fixed main effects for gender (using two dummy variables for female and other; male as reference category), age (grand-mean centered), and education (grand-mean centered). The robustness check repeated the non-significant effects.

Controlling for gender, age, and education, the attitudinal dissimilarity × writing task condition × ideological extremity interaction effect remained non-significant $b = 0.01$, $SE = 0.02$, $t(466) = 0.34$, $p = .369$ (one-tailed). The predicted indirect effects remained non-significant as well, Sobel test for mediation via strength of moral convictions: $z = -0.22$, $p = .585$ (one-tailed); Sobel test for mediation via intolerance of political ambiguity: $z = 1.21$, $p = .114$ (one-tailed).

In short, across all of the confirmatory analyses we did not find support for our hypotheses. Although attitudinal dissimilarity was robustly associated with prejudice, replicating decades of past work on the topic, this link was not reduced in size for political moderates by puncturing the illusion of explanatory depth.

Exploratory Analyses

Change in Understanding – Potential Moderators. We conducted exploratory analyses to test whether our manipulation worked only for some subgroups or across the board. Therefore, we added main effects and two-way interactions effects with the writing task condition for ideological extremity, strength of moral conviction, and intolerance of political ambiguity to the model. However, all of these effects were non-significant ($p > .098$). That is, we did not find evidence that the change in understanding ratings depended on any of our proposed moderators.

Simplifying Main Model. We conducted additional exploratory analyses in which we removed all non-significants effects from our main model. All interaction effects involving the writing task condition and its main effect were non-significant (all ps $> .115$). Thus, the final model consisted of two main effects for attitudinal dissimilarity and ideological extremity and their interaction effect.

This interaction effect was significant, $b = 0.03$, $SE = 0.01$, $t(469) = 3.94$, $p < .001$. For moderates, the effect of attitudinal dissimilarity on prejudice was strong and significant, $b = 0.24$,

SE = 0.03, *t*(482) = 9.18, *p* < .001. For extremists, the effect of attitudinal dissimilarity on prejudice was even stronger, *b* = 0.41, *SE* = 0.02, *t*(455) = 16.31, *p* < .001.

Simplified Mediation Model. In addition, we conducted exploratory analyses in which we removed all non-significants effects from our mediation model. The attitudinal dissimilarity × writing task condition × ideological extremity interaction effect, the attitudinal dissimilarity × writing task condition × strength of moral conviction interaction effect, the writing task condition × ideological extremity interaction effect, and the writing task condition × strength of moral conviction interaction effect were non-significant (all *p*s > .708) and removed from the model. Thus, the final model consisted of five main effects (attitudinal dissimilarity, writing task condition, ideological extremity, strength of moral conviction, and intolerance of political ambiguity), five two-way interaction effects (attitudinal dissimilarity × writing task condition, attitudinal dissimilarity × ideological extremity, attitudinal dissimilarity × strength of moral conviction, attitudinal dissimilarity × intolerance of political ambiguity, and writing task condition × intolerance of political ambiguity) and one three-way interaction effect (attitudinal dissimilarity × writing task condition × intolerance of political ambiguity).

The attitudinal dissimilarity × writing task condition × intolerance of political ambiguity interaction effect remained significant, *b* = 0.39, *SE* = 0.18, *t*(460) = 2.09, *p* = .019 (one-tailed). The attitudinal dissimilarity × ideological extremity also remained significant, *b* = 0.03, *SE* = 0.01, *t*(464) = 3.19, *p* = .002. In addition, the attitudinal dissimilarity × strength of moral conviction interaction effect was also significant, *b* = 0.27, *SE* = 0.08, *t*(476) = 3.23, *p* = .001. For participants with relatively weak moral convictions, the effect of attitudinal dissimilarity on prejudice was strong and significant, *b* = 0.20, *SE* = 0.03, *t*(485) = 7.17, *p* < .001. For participants with relatively strong moral convictions, the effect of attitudinal dissimilarity on prejudice was even stronger, *b* = 0.30, *SE* = 0.03, *t*(469) = 9.25, *p* < .001.

Moderator: Ideological Type. For our confirmatory main analyses, we did not include the attitudinal dissimilarity × writing task condition × ideological type interaction effect because it did not meet our pre-registered criterion of statistical significance ($p < .05$). However, considering that we found tentative evidence for a moderating effect of ideological type in our earlier study and as the effect was close to significance in the analysis leading up to our main model ($p = .067$), we decided to explore it in more detail. Therefore, we added the attitudinal dissimilarity × writing task condition × ideological type interaction effect and the corresponding lower-order main and interaction effects to the simplified main model.

The attitudinal dissimilarity × writing task condition × ideological type interaction effect was not significant, $b = -0.06$, $SE = 0.04$, $t(460) = -1.75$, $p = .082$. Similarly, the attitudinal dissimilarity × writing task condition interaction effect was not significant for both conservatives, $b = 0.07$, $SE = 0.06$, $t(462) = 1.18$, $p = .239$, and liberals, $b = -0.06$, $SE = 0.04$, $t(463) = -1.55$, $p = .121$. That is, puncturing the illusion of explanatory depth did not appear to reduce the relationship between attitudinal dissimilarity and prejudice for neither liberals nor conservatives.

Discussion

The relationship between attitudinal dissimilarity and prejudice is one of the most robust relationships in social psychology (Brandt et al., 2014, 2015; Byrne, 1969; Byrne & Nelson, 1965). In this research, we replicated this effect. However, we also hypothesised and tested the effectiveness of puncturing the illusion of explanatory depth for reducing the strength of, or even breaking, this relationship. In an initial study, we found support for this hypothesis for political moderates but not for political extremists. In the main study, we aimed to replicate this effect and tested whether it would be mediated by two potential psychological characteristics of moderates: relatively weak moral convictions and relatively high tolerance of political ambiguity.

The results of the main study did *not* provide support for the idea that either political moderates or extremists who have had their illusion of explanatory depth punctionured become more tolerant of other people perceived as being attitudinally dissimilar from themselves. Similarly, we found no evidence for such an effect among people with relatively strong or weak moral convictions or people who identify as politically conservative or liberal.

The result for intolerance of political ambiguity needs additional elaboration. We found a significant moderating effect of intolerance of political ambiguity for the attitudinal dissimilarity × writing task condition in the predicted direction. However, the predicted attitudinal dissimilarity × writing task condition interaction effect for participants who are relatively tolerant of political ambuity was not significant (i.e. the simple effect was not significant). Notably, the magnitude of this effect was substantial, but the effect was unreliable. This issue may stem from the fact the items from the adapted scale that we used to measure intolerance of political ambiguity were considerably less reliable than we expected.

Although a failure to reject the null hypothesis is difficult to interpret, our conservative power analyses suggested that we had at least a 80% chance of finding a statistically significant three-way interaction effect of small size and a near 100% chance of finding a statistically significant three-way interaction effect of a medium size, assuming such an effect exists. Thus, we cannot rule out that our study was a false negative and puncturing of the illusion of explanatory depth *does* reduce the relationship between attitudinal dissimilarity and prejudice for political moderates or people with weak moral convictions. However, if our power analyses are correct and restricted to the population of MTurkers, it is highly unlikely that such effects are of medium or strong size. At best, these effects are small, which may limit their practical utility for concrete real-world interventions.

Digging Deeper: Which Processes Could Be at Work

An important starting point for this discussion is that we did find that participants in the mechanism explanation condition, but not in the reason generation condition, showed statistically significant drops in estimates of what they believed they understood. That is, the manipulation worked as expected. Therefore, the reason the predicted effects were not found must lay somewhere else.

Our theoretical approach hypothesized a reduction-effect of the puncturing of explanatory depth manipulation for political moderates. However, it could be possible that puncturing the illusion of explanatory depth activates multiple processes that simultaneously strengthen and weaken the association between attitudinal dissimilarity and prejudice. For example, social identity and realistic group conflict theories suggest a relationship between ingroup-identification and outgroup hostility under conditions of intergroup threat (Duckitt & Mphuthing, 1998). If participants realize that the basis of their preference for their ingroup over outgroups is less warranted than previously thought, they may react with increased hostility towards the outgroup (cf. Heine, Proulx, & Vohs, 2006). Such a process may cancel out the effect that we hypothesized. Including measures of ingroup identification and perceived intergroup threat in future studies could help to disentangle these potentially opposing effects.

In addition, the prejudice literature suggests alternative subgroups for which the proposed intervention might be more of less effective. For example, the need for cognitive closure is closely related to intolerance of political ambiguity. It is defined as “a desire for an answer on a given topic, any answer, . . . compared to confusion and ambiguity" (Webster & Kruglanski, 1994, p. 1049). This desire has been identified as the general motivated cognitive style underlying prejudice (Roets & Van Hiel, 2011a). When we proposed the mediators for our study, we argued that intolerance of political ambiguity might be a better choice than need for cognitive closure as it is specific to the political domain. However, our results for intolerance of political

ambiguity must be treated with caution due to the low reliability coefficient for the scale that we adapted from previous research (McClosky & Chong, 1985). Validated scales to measure the more general need for closure (e.g., Roets & Van Hiel, 2011b) propose a way to both addressing our reliability issue and extending our results to a more general style of avoiding versus embracing ambiguity.

A second possibility for future research is to investigate honesty-humility as a potential moderator. Honesty-humility is a dimension of the HEXACO model of personalit structure. It "reflects an orientation towards fairness and sincerity in social relations versus the tendency to manipulate and use people for whatever one can get from them" (Sibley, Harding, Perry, Asbrock, & Duckitt, 2010, p. 517). Past research has identified a complex role for honesty-humility as a contributor to prejudice (Bergh & Akrami, 2016; Sibley et al., 2010; Stuermer et al., 2013). However, if people with a honest and humble personality are particularly concerned about treating other social groups fairly, they might be especially likely to become more tolerant of dissimilar others when realizing that they overestimated their understanding of important policies.

One common assumption throughout this "digging deeper" discussion is that puncturing the illusion of explanatory depth can reduce the link between dissimilarity and prejudice for at least some people through some mechanisms. However, it is important to be clear that this an assumption for the sake of discussion. It is also possible that puncturing the illusion of explanatory depth does not reduce the association between dissimilarity and prejudice. Future researchers should procede with caution.

Limitations and Future Directions

Our research presents several limitations. First, although Amazon's Mechanical Turk samples are of similar (if not better) quality than other convenience samples like college students,

it is not representative of any population (Berinsky, Huber, & Lenz, 2012; Buhrmester, Kwang, & Gosling, 2011; Huff & Tingley, 2015; Paolacci & Chandler, 2014). Our sample consisted of American MTurkers because our priority was on having enough power to detect the effect we had hypothesized. Considering the large sample size we pursued, MTurk provides a platform to conduct our study within a reasonable cost and time frame. However, as MTurk is a US based platform that is usually liberally skewed (i.e. there are more liberal than conservative MTurkers; Berinsky et al., 2012; Huff & Tingley, 2015), our findings are limited to this sample. Future research could examine whether puncturing the illusion of explanatory depth successfully reduces the relationship between attitudinal dissimilarity and prejudice in countries with a multipolar or with less polarized political scene than in the US.

Second, in both the initial and the main study, our study suffered from high dropout rates. In addition, the dropout rate was significantly higher in the mechanism explanation condition than in the reason generation condition. Similar differences were also observed in both of Fernbach et al.'s (2013) studies. This finding might be problematic, because if certain types of people are more likely to drop out in one condition than in another, researchers can no longer assume that the principle of random assignment ensures no non-random a priori differences between the experimental conditions (Zhou & Fishnbach, 2016). Future methodological research should aim to modify the mechanism generation intervention in order to counter these high dropout rates.

Finally, another direction for future research is to examine whether the consequences of realizing that one knows less about the mechanisms of policy interventions than one initially assumes change over time. Other research has suggested that attitudinal dissimilarity becomes more important in influencing group cohesiveness over time as group members accumulate more information about each other (Harrison, Price, & Bell, 1998). We propose that reducing the effect

of dissimilarity on prejudice may similarly take time. While realizing that one knows less than one thought might be painful or unpleasant in the first moment, and one might want to devalue other, dissimilar groups, people may, in the long run, question the basis for their prejudice and learn to be more tolerant towards others from their illusion of explanatory depth. Thus, studies could compare the short-term and the long-term effects of puncturing the illusion of explanatory depth.

References

Amodio, D. M. (2014). The neuroscience of prejudice and stereotyping. *Nature Reviews Neuroscience, 15*, 670-682.

Aramovich, N. P., Lytle, B. L., & Skitka, L. J. (2012). Opposing torture: Moral conviction and resistance to majority influence. *Social Influence, 7*(1), 21-34.

Bassili, J. N. (1996). Meta-judgmental versus operative indexes of psychological attributes: The case of measures of attitude strength. *Journal of Personality and Social Psychology, 71*(4), 637-653.

Bergh, R., & Akrami, N. (2016). Are non-agreeable individuals prejudiced? Comparing different conceptualizations of agreeableness. *Personality and Individual Differences, 101*, 153-159.

Berinsky, A. J., Huber, G. A., & Lenz, G. S. (2012). Evaluating online labor markets for experimental research: Amazon. com's Mechanical Turk. *Political Analysis, 20*(3), 351-368.

Boyer, P., Firat, R., & van Leeuwen, F. (2015). Safety, threat, and stress in intergroup relations: A coalitional index model. *Perspectives on Psychological Science, 10*(4), 434-450.

Brandt, M. J. (2017). Predicting ideological prejudice. *Psychological Science, 28*(6), 713-722.

Brandt, M. J., Chambers, J. R., Crawford, J. T., Wetherell, G., & Reyna, C. (2015). Bounded openness: The effect of openness to experience on intolerance is moderated by target group conventionality. *Journal of Personality and Social Psychology, 109*(3), 549-568.

Brandt, M. J., Evans, A. M., & Crawford, J. T. (2015). The unthinking or confident extremist? Political extremists are more likely than moderates to reject experimenter-generated anchors. *Psychological Science, 26*(2), 189-202.

Brandt, M. J., & Proulx, T. (2016). Conceptual creep as a human (and scientific) goal. *Psychological Inquiry, 27*(1), 18–23.

Brandt, M. J., Reyna, C., Chambers, J. R., Crawford, J. T., & Wetherell, G. (2014). The ideological-conflict hypothesis intolerance among both liberals and conservatives. *Current Directions in Psychological Science, 23*(1), 27-34.

Brown, R. (2010). *Prejudice: Its social psychology* (2nd ed.). Malden, MA: Wiley-Blackwell.

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science, 6*(1), 3-5.

Byrne, D. (1969). Attitudes and attraction. *Advances in Experimental Social Psychology, 4*, 35–89.

Byrne, D., & Nelson, D. (1965). Attraction as a linear function of proportion of positive reinforcements. *Journal of Personality and Social Psychology, 1*, 659.

Cameron, L., Rutland, A., Brown, R., & Douch, R. (2006). Changing children's intergroup attitudes toward refugees: Testing different models of extended contact. *Child Development, 77*(5), 1208-1219.

Chambers, J. R., & Melnyk, D. (2006). Why do I hate thee? Conflict misperceptions and intergroup mistrust. *Personality and Social Psychology Bulletin, 32*(10), 1295–1311.

Chambers, J. R., Schlenker, B. R., & Collisson, B. (2013). Ideology and prejudice: The role of value conflicts. *Psychological Science, 24*(2), 140-149.

Clinton, B. (2014, November 20). We only have one remaining bigotry: We don’t want to be around anybody who disagrees with us. *New Republic*. Retrieved from https://goo.gl/oRtgud

Colombo, M., Bucher, L., & Inbar, Y. (2016). Explanatory judgment, moral offense and value-free science. *Review of Philosophy and Psychology, 7*(4), 743-763.

Cox, W. T., & Devine, P. G. (2013). Stereotyping to infer group membership creates plausible deniability for prejudice-based aggression. *Psychological Science, 25*(2), 340-348.

Crandall, C. S., Eshleman, A., & O'Brien, L. (2002). Social norms and the expression and suppression of prejudice: the struggle for internalization. *Journal of Personality and Social Psychology, 82*(3), 359-378.

Crandall, C. S., Ferguson, M. A., & Bahns, A. J. (2013). When we see prejudice: The normative window and social change. In C. Stangor & C. S. Crandall (Eds.), *Frontiers in stereotyping and prejudice* (pp. 53–70). New York, NY: Psychology Press.

Crawford, J. T. (2014). Ideological symmetries and asymmetries in political intolerance and prejudice toward political activist groups. *Journal of Experimental Social Psychology, 55*, 284-298.

Crawford, J. T., & Pilanski, J. M. (2014). Political intolerance, right and left. *Political Psychology, 35*(6), 841-851.

Dovidio, J. F., & Gaertner, S. L. (1999). Reducing prejudice: Combating intergroup biases. *Current Directions in Psychological Science, 8*(4), 101-105.

Duckitt, J., & Mphuthing, T. (1998). Group identification and intergroup attitudes: a longitudinal analysis in South Africa. *Journal of Personality and Social Psychology, 74*(1), 80-85.

Fernbach, P. M., Rogers, T., Fox, C. R., & Sloman, S. A. (2013). Political extremism is supported by an illusion of understanding. *Psychological Science, 24*(6), 939-946.

Gaertner, S. L., Mann, J. A., Dovidio, J. F., Murrell, A. J., & Pomare, M. (1990). How does cooperation reduce intergroup bias? *Journal of Personality and Social Psychology, 59*(4), 692-704.

Goodwin, G. P., Piazza, J., & Rozin, P. (2014). Moral character predominates in person perception and evaluation. *Journal of Personality and Social Psychology, 106*(1), 148-168.

Graham, J., Haidt, J., & Nosek, B. A. (2009). Liberals and conservatives rely on different sets of moral foundations. *Journal of Personality and Social Psychology, 96*(5) 1029-1046.

Harrison, D. A., Price, K. H., & Bell, M. P. (1998). Beyond relational demography: Time and the effects of surface-and deep-level diversity on work group cohesion. *Academy of Management Journal, 41*(1), 96-107.

Heine, S. J., Proulx, T., & Vohs, K. D. (2006). The meaning maintenance model: On the coherence of social motivations. *Personality and Social Psychology Review, 10*(2), 88-110.

Hornsey, M. J., & Hogg, M. A. (2000). Intergroup similarity and subgroup relations: Some implications for assimilation. *Personality and Social Psychology Bulletin, 26*(8), 948-958.

Huff, C., & Tingley, D. (2015). "Who are these people?" Evaluating the demographic characteristics and political preferences of MTurk survey respondents. *Research & Politics, 2*(3), 1-12.

Koch, A., Imhoff, R., Dotsch, R., Unkelbach, C., & Alves, H. (2016). The ABC of stereotypes about groups: Agency/socioeconomic success, conservative–progressive beliefs, and communion. *Journal of Personality and Social Psychology, 110*(5), 675-709.

Kurzban, R., Tooby, J., & Cosmides, L. (2001). Can race be erased? Coalitional computation and social categorization. *Proceedings of the National Academy of Sciences, 98*(26), 15387-15392.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2016). *lmerTest: Tests in Linear Mixed Effects Models*. Retrieved from https://cran.r-project.org/web/packages/lmerTest/index.html

Lammers, J., Koch, A., Conway, P., & Brandt, M. J. (2016). The political domain appears simpler to the politically extreme than to political moderates. *Social Psychological and Personality Science, 8*(6), 612-622.

McClosky, H., & Chong, D. (1985). Similarities and differences between left-wing and right-wing radicals. *British Journal of Political Science, 15*(3), 329-363.

Paolacci, G., & Chandler, J. (2014). Inside the Turk: Understanding Mechanical Turk as a participant pool. *Current Directions in Psychological Science, 23*(3), 184-188.

Park, B., & Judd, C. M. (2005). Rethinking the link between categorization and prejudice within the social cognition perspective. *Personality and Social Psychology Review, 9*(2), 108-130.

Pettigrew, T. F., & Tropp, L. R. (2006). A meta-analytic test of intergroup contact theory. *Journal of Personality and Social Psychology, 90*(5), 751-783.

Pietraszewski, D., Curry, O. S., Petersen, M. B., Cosmides, L., & Tooby, J. (2015). Constituents of political cognition: Race, party politics, and the alliance detection system. *Cognition, 140*, 24-39.

Preacher, K. J., & Leonardelli, G. J. (2001). *Calculation for the Sobel test: An interactive calculation tool for mediation tests*. Retrieved from http://quantpsy.org/sobel/sobel.htm

Richard, F. D., Bond Jr., C. F., & Stokes-Zoota, J. J. (2003). One hundred years of social psychology quantitatively described. *Review of General Psychology, 7*(4), 331–363.

Roets, A., & Van Hiel, A. (2011a). Allport's prejudiced personality today: Need for closure as the motivated cognitive basis of prejudice. *Current Directions in Psychological Science, 20*(6), 349-354.

Roets, A., & Van Hiel, A. (2011b). Item selection and validation of a brief, 15-item version of the Need for Closure Scale. *Personality and Individual Differences, 50*(1), 90-94.

Rozenblit, L., & Keil, F. (2002). The misunderstood limits of folk science: An illusion of explanatory depth. *Cognitive Science, 26*(5), 521-562.

Rucker, D. D., Preacher, K. J., Tormala, Z. L., & Petty, R. E. (2011). Mediation analysis in social psychology: Current practices and new recommendations. *Social and Personality Psychology Compass, 5*(6), 359-371.

Ryan, T. J. (2014). Reconsidering moral issues in politics. *The Journal of Politics, 76*(2), 380-397.

Schaller, M., Boyd, C., Yohannes, J., & O'Brien, M. (1995). The prejudiced personality revisited: Personal need for structure and formation of erroneous group stereotypes. *Journal of Personality and Social Psychology, 68*(3), 544-555.

Sibley, C. G., Harding, J. F., Perry, R., Asbrock, F., & Duckitt, J. (2010). Personality and prejudice: Extension to the HEXACO personality model. *European Journal of Personality, 24*(6), 515-534.

Skitka, L. J., Bauman, C. W., & Lytle, B. L. (2009). Limits on legitimacy: Moral and religious convictions as constraints on deference to authority. *Journal of Personality and Social Psychology, 97*(4), 567-578.

Skitka, L. J., Bauman, C. W., & Sargis, E. G. (2005). Moral conviction: Another contributor to attitude strength or something more? *Journal of Personality and Social Psychology, 88*(6), 895-917.

Stangor, C. (2009). The study of stereotyping, prejudice, and discrimination within social psychology: A quick history of theory and research. In T. D. Nelson (Ed.), *Handbook of prejudice, stereotyping, and discrimination* (pp. 1–22). New York, NY: Psychology Press.

Stathi, S., Cameron, L., Hartley, B., & Bradford, S. (2014). Imagined contact as a prejudice-reduction intervention in schools: The underlying role of similarity and attitudes. *Journal of Applied Social Psychology, 44*(8), 536-546.

Stuermer, S., Benbow, A. E., Siem, B., Barth, M., Bodansky, A. N., & Lotz-Schmitt, K. (2013). Psychological foundations of xenophilia: The role of major personality traits in predicting favorable attitudes toward cross-cultural contact and exploration. *Journal of Personality and Social Psychology, 105*(5), 832-851.

Tetlock, P. E. (1984). Cognitive style and political belief systems in the British House of Commons. *Journal of Personality and Social Psychology, 46*(2), 365-375.

Toner, K., Leary, M. R., Asher, M. W., & Jongman-Sereno, K. P. (2013). Feeling superior is a bipartisan issue: Extremity (not direction) of political views predicts perceived belief superiority. *Psychological Science, 24*(12), 2454-2462.

Tormala, Z. L., & Petty, R. E. (2002). What doesn't kill me makes me stronger: The effects of resisting persuasion on attitude certainty. *Journal of Personality and Social Psychology, 83*(6), 1298–1313.

Webster, D. M., & Kruglanski, A. W. (1994). Individual differences in need for cognitive closure. *Journal of Personality and Social Psychology, 67*(6), 1049-1062.

Westfall, J. (2016). *PANGEA: Power ANalysis for GEneral Anova designs* (Working paper). Retrieved from http://jakewestfall.org/publications/pangea.pdf

Wisneski, D. C., Lytle, B. L., & Skitka, L. J. (2009). Gut reactions: Moral conviction, religiosity, and trust in authority. *Psychological Science, 20*(9), 1059-1063.

Wynn, K. (2016). Origins of value conflict: Babies do not agree to disagree. *Trends in Cognitive Sciences, 20*(1), 3–5.

Zhang, Z., Zyphur, M. J., & Preacher, K. J. (2009). Testing multilevel mediation using hierarchical linear models: Problems and solutions. *Organizational Research Methods, 12*(4), 695-719.

Zhou, H., & Fishbach, A. (2016). The pitfall of experimenting on the web: How unattended selective attrition leads to surprising (yet false) research conclusions. *Journal of Personality and Social Psychology, 111*(4), 493-504.

Appendix

List of Social Groups

Whites	Democrats	Blacks	Poor	Middle class
Asians	Rich	Atheists	Republicans	Christians
Liberals	Conservatives	Nerds	Students	Athletes
Jews	Hispanics	Women	Artists	Musicians
Teenagers	Muslims	Politicians	Catholics	Men
Jocks	Preps	Elderly	Business people	White-collar

Disclosure Statement

All authors have no financial interest or benefit arising from the direct applications of their research.